



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



Towards the Contextualization of Provenance Information - Experiences with CIDOC-CRM and WissKI

Stefanie Rühle und Jürgen Dönitz
(SUB Göttingen)



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN



Overview

- About ASCH
- Modelling Provenance Information
- Test
 - WissKI
 - CIDOC-CRM
- Open Issues

About the Project



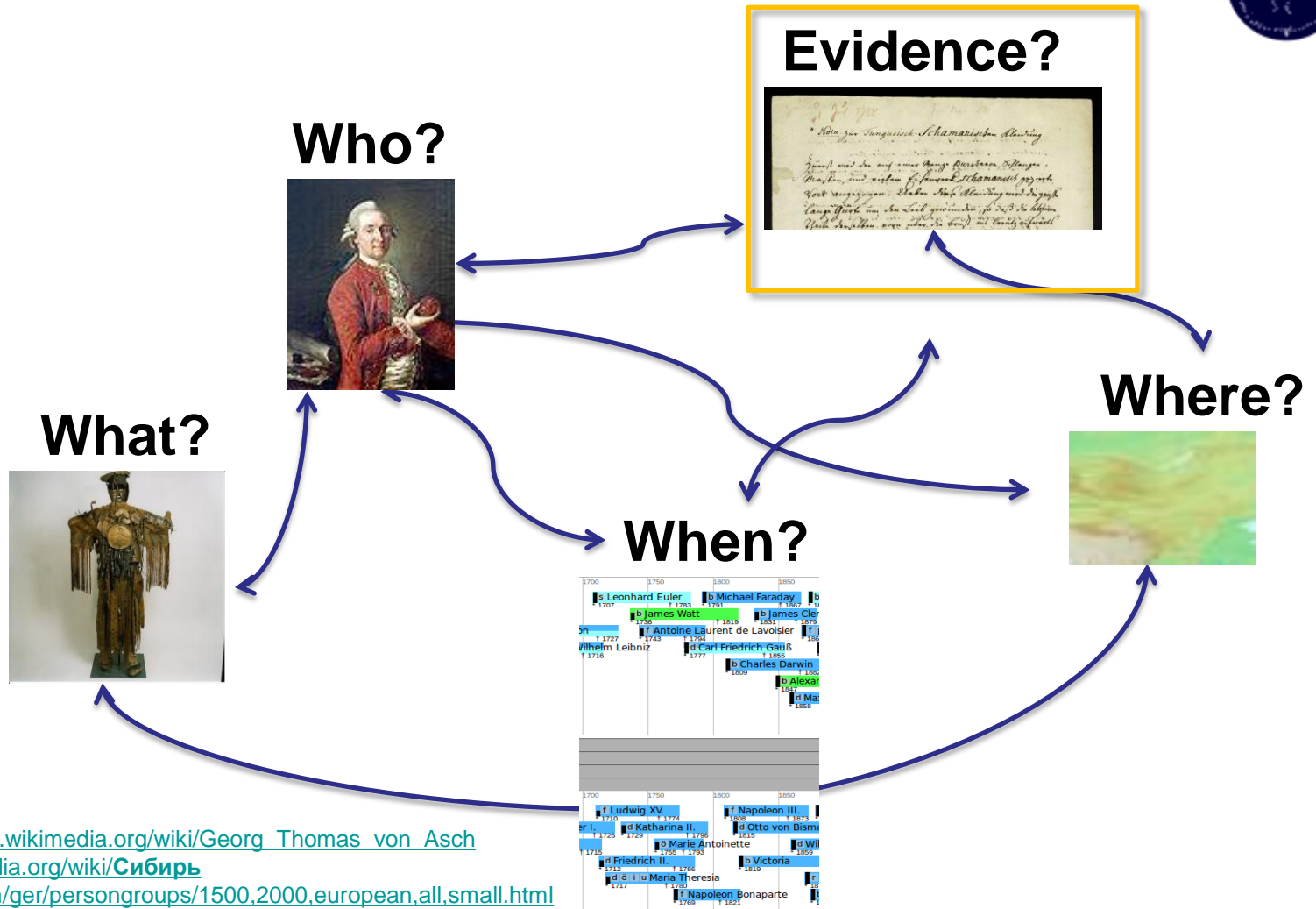
- Purpose and Scope:
 - developing a metadata model for contextualising heterogeneous objects from collections
 - with special focus on provenance information
- Funding
 - by Deutsche Forschungsgemeinschaft
 - 01.09. 2014 – 31.08. 2017
- Project Lead:
 - Göttingen State and University Library
 - Institute of Social and Cultural Anthropology at the University of Göttingen
- Contact
 - Susanne Al-Eryani (mailto: al-eryani(a)sub.uni-goettingen.de)
 - Gudrun Bucher (mailto: gudrun.bucher(a)sub.uni-goettingen.de)
 - Jürgen Dönitz (mailto: doenitz(a)sub.uni-goettingen.de)



Description of Provenance

- „... descriptions of the entities and activities involved in producing and delivering or otherwise influencing a given object.“
Source: <https://www.w3.org/TR/prov-overview/>
- Description of the events in the lifecycle of an object
 - LIDO (Lightweight Information Describing Objects)
 - CIDOC-CRM (Conceptual Reference Model of the International Committee for Documentation of the International Council of Museums)
 - W3C PROV (a specification to express provenance records)
- Involved entities: what, who, when, where
- Evidence verifying statements

Contextualization of Provenance Information

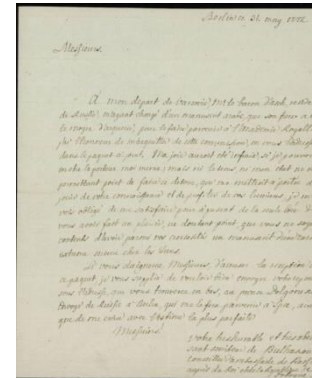


Source: http://commons.wikimedia.org/wiki/Georg_Thomas_von_Asch
<http://commons.wikimedia.org/wiki/Сибирь>
<http://www.vistorica.com/ger/persongroups/1500,2000,european,all,small.html>



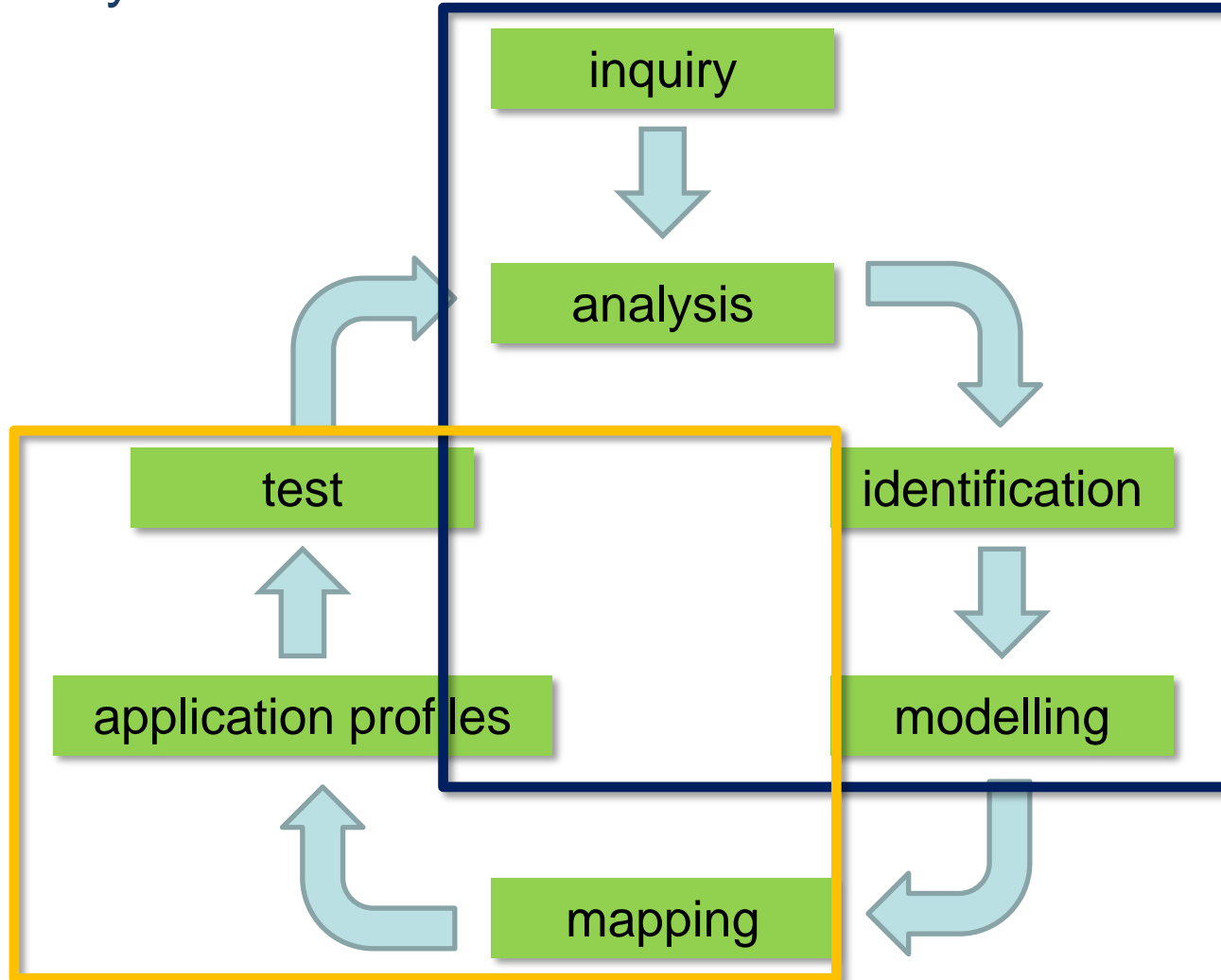
Why Provenance “von Asch”?

- The items
 - books, maps, coins, rocks, engravings, skulls, stuffed animals, seed, ...
- The origin
 - Sibiria, Russian America, Black Sea, Turkey, ...
- The evidence
 - letters, inventories, lists, label, ...





7 Steps to your Metadata Model





Modeling the Use Cases

- UC 1 Information about resources
- UC 2 Identification of resources
- **UC 3 Information about the history/lifecycle of resources**
- UC 4 Information about change of use and reception of resources
- **UC 5 Proof of information by evidence**
- **UC 6 Reliability of statements**
- UC 7 Access to resources
- UC 8 Reuse of data

(http://asch.wiki.gwdg.de/index.php/Use_Cases)

Identification of relevant classes and relations between these classes



- Classes

- resource
- metadata set
- item
- **evidence**
- **event**
- time
- agent
- place
- digital representation
- collection
- statement
- holding
- concept

Event

:Event

Type of Term: Class

Definition: An activity in the lifecycle of a resource.

:hasEvidence

Type of term: rdfs:Property

Definition: relates an event with resources proving the existence of the event

Domain: Event

Range: Evidence

Occurrence: 0-n

Subproperty of:

Evidence

:Evidence

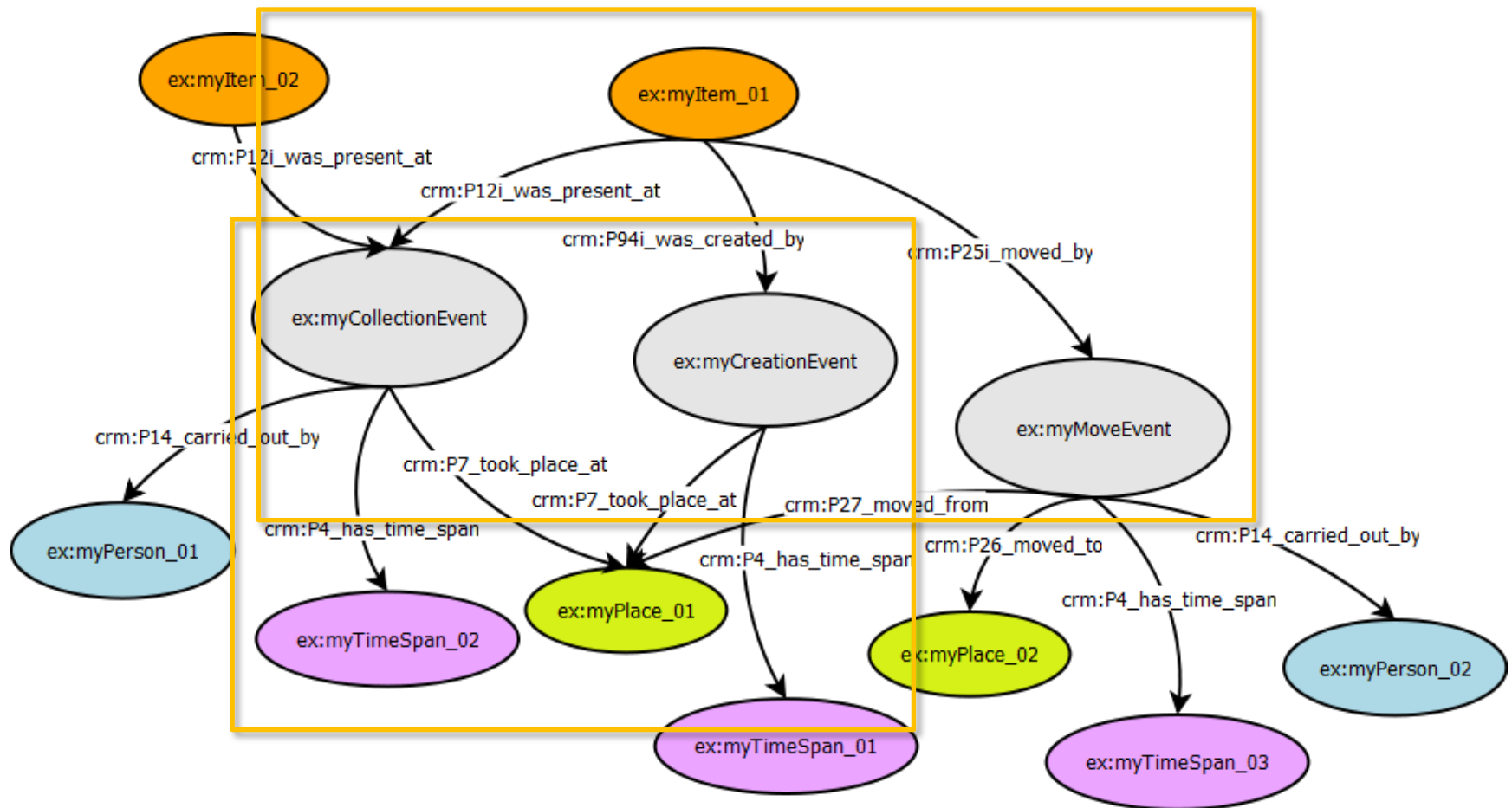
Type of Term: Class

Definition: A resource proving the reliability of a statement about a resource

Subclass of:

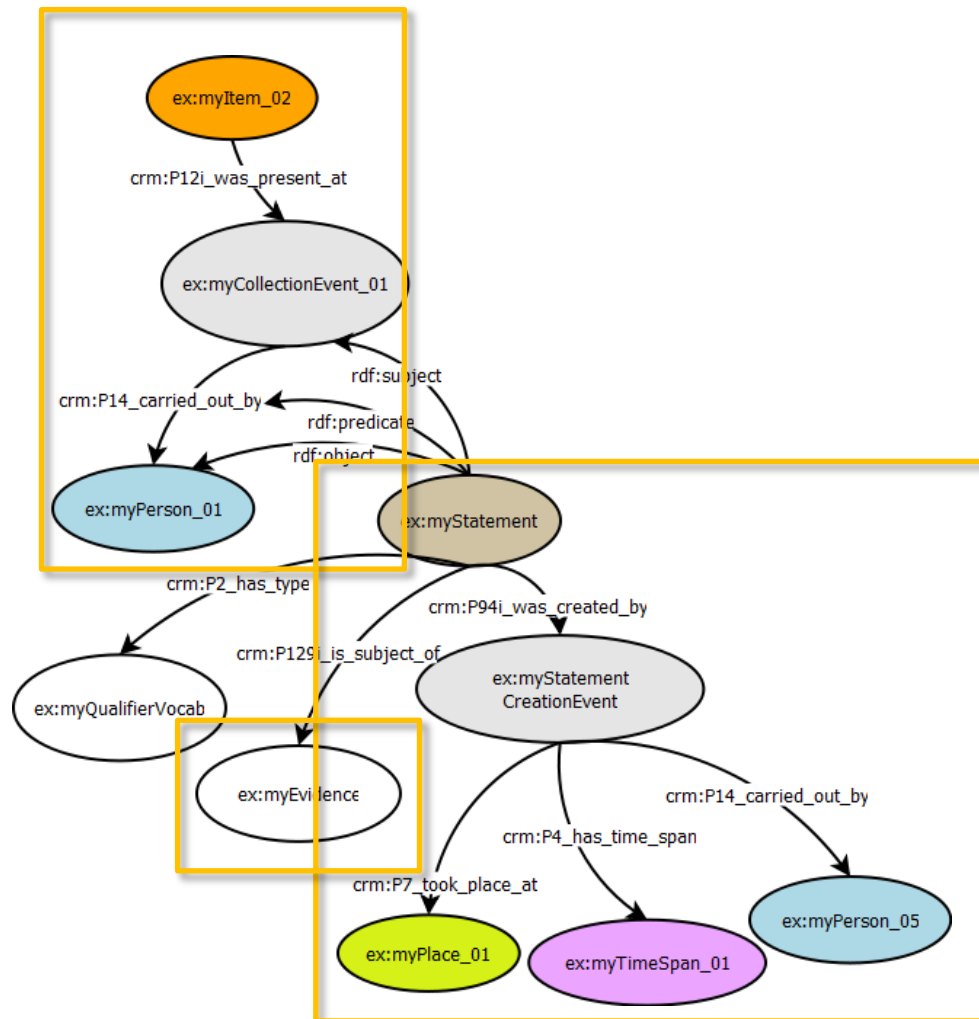


Event-based Description of the Lifecycle





Description of Provenance Statements



Test



- Test the model in different software environments:

- WissKI → CIDOC-CRM AP
- Semantic Media Wiki
- OntoWiki
- TextGrid
- Material Evidence in Incunabula (MEI)



OntoWiki





Experiences installing and operating WissKI

- Task in the project:
 - Evaluate annotation systems
 - Installation, setup and annotation of a few test data samples
 - No development or debugging
- Requirements of WissKI (v1) (PHP, MySQL, Drupal)
 - Components not up-to-date or even at end of life
 - Security issues prohibit public access to the web interface
- Installation (only) with the support of the developers
 - Image module is required to produce URIs for new nodes; it is only recommended

Using (setting up) WissKI

- (subjective) maturity of the WissKI
 - Project with limited resources
 - Limited number of use cases / workflows in mind?
- Is the sequence of steps important?
 - After adding a path, the page with drupal modules is broken
 - Path groups can't be changed after the paths are changed
 - Sometimes SQL errors
- The search function is not working (empty list)
- The navigation is broken after adding a controlled vocabulary
- Most of the documentation stops with the creation of the paths
 - By default paths are disabled -> strange behavior while creating content





Wishes for current / further development

- Open process
 - Github, Issue tracker
- Include as many projects as possible to gather requirements to test workflows
- Long term support, independent of funding time
- Easy installation, e.g. with docker (one click installer)
- Documentation for setting up a project





Requirements from the Asch Project for WissKI

- Virtual collections
- Import / Export from and to various formats
 - Single datasets, selected ones, all
- Simple backup / dump
- Change history & Undo
- Controlled Vocabularies:
 - Immutable CVs, user extendible
- Objects and relations have to be identifiable, IDs should be unique also after deletion of the entity.
- Relations need additional information, e.g. valid time, references, author
- One or more images for each node
- (Optional) validation of user input
- User authorization and different roles



CRM and ASCH Model

- Suitable to describe provenance information
- Properties and entities don't meet all requirements
 - E.g. P49 has former or current keeper and P50 has current keeper but no has former keeper
- „Simple statements“ may become rather complex
 - E73 Information Object P106 is composed of E33 Linguistic Object P72 has language E56 Language P1 is identified by E41 Appellation
 - or
 - E73 Information Object P128i is carried by E84 Information Carrier P46i forms part of E78 Collection
- The use of statements about statements is limited.

=> merge with other models



Open Issues

- University Collections -> a stable, reliable, comprehensive database
- CIDOC CRM
 - Suitable to contextualize data but
 - as backbone of the metadata editor – does that make sense?
 - as exchange format – CRM compliant import and export?
 - what to do with administrative data -> e.g. Spectrum Workflow, management of digital representations?
- WissKI and University Collections
 - Reliability and sustainability of the system?
 - Documentation?
 - Wisski without CRM = other ontologies?



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



Any questions?

Contact:

Susanne Al-Eryani (mailto: al-eryani(a)sub.uni-goettingen.de)

Gudrun Bucher (mailto: gudrun.bucher(a)sub.uni-goettingen.de)

Jürgen Dönitz (mailto: doenitz(a)sub.uni-goettingen.de)

Stefanie Rühle (mailto: sruehle(a)sub.uni-goettingen.de)

More information about ASCH at:

http://asch.wiki.gwdg.de/index.php/Main_Page



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN